

Small Stress Response Proteins in *Escherichia coli*: Proteins Missed by Classical Proteomic Studies

MATTHEW R. HEMM,[†] BRIAN J. PAUL,[‡] JUAN MIRANDA-RIOS,[§] AIXIA ZHANG,
NIMA SOLTANZAD,^{1,1} AND GISELA STORZ*

*Cell Biology and Metabolism Program, Eunice Kennedy Shriver National Institute of Child
Health and Human Development, Bethesda, MD 20892-5430*

* Corresponding author. Mailing address: NIH, Building 18T, Room 101, 18 Library Dr.,
MSC 5430, Bethesda, MD 20892-5430. Phone: (301) 402-0968. Fax: (301) 402-0078. E-mail:
storz@helix.nih.gov.

[†] Present address: Department of Biology, Towson University, Towson, MD
21252.

[‡] Present address: DuPont Central Research and Development, Wilmington,
DE 19880.

22 § Present address: Unidad de Genética de la Nutrición, Instituto de
23 Investigaciones Biomédicas, Universidad Nacional Autónoma de México, and
24 Instituto Nacional de Pediatría, México, D.F. 04510, México.

25 ^{1 1} Present address: Weill Medical College of Cornell University, New York,
26 NY 10021.

27

28

29 Running title: Stress-induced small proteins

30 Keywords: Zur, CRP, heat shock, Sequential Peptide Affinity tag

31 Proteins of fifty or fewer amino acids are poorly characterized in all organisms.
32 The corresponding genes are challenging to reliably annotate, and it is difficult to purify
33 and characterize the small protein products. Due to these technical limitations, little is
34 known about the abundance of small proteins, not to mention their biological functions. To
35 begin to characterize these small proteins in *Escherichia coli*, we assayed their
36 accumulation under a variety of growth conditions and after exposure to stress. We found
37 that many small proteins accumulate under specific growth conditions or are stress
38 induced. For some genes, the observed changes in protein levels were consistent with
39 known transcriptional regulation, such as ArcA activation of the operons encoding *yccB*
40 and *ybgT*. However we also identified novel regulation, such as Zur repression of *ykgMO*,
41 CRP repression of *azuC*, and CRP activation of *ykgR*. The levels of eleven small proteins
42 increase after heat shock and induction of at least one of these, YobF, occurs at a post-
43 transcriptional level. These results show that small proteins are an overlooked subset of
44 stress response proteins in *E. coli*, and provide information that will be valuable for
45 determining the functions of these proteins.

46 A challenge of whole-proteome studies of any biological system is the identification and
47 characterization of small proteins, herein defined as those having 50 or fewer amino acids.
48 These small proteins are difficult to detect using standard biochemical techniques. For the
49 analysis of proteins in complex lysates, many studies have employed two-dimensional gel
50 electrophoresis, a technique that is biased toward abundant proteins of standard size (30-200
51 kDa) (21, 29). Proteins that are present at low levels or have extremely large or small molecular
52 weights are usually missed in these experiments (11, 21, 29, 38). Recently, methods for using
53 mass spectroscopy to analyze proteins from crude mixtures have been developed, and can
54 provide higher sensitivity and resolution (21). However, small proteins are still difficult to
55 identify using these methods. This is in part due to the fact that fewer peptides are produced
56 from peptidic degradation of small proteins, making it more difficult to identify these proteins
57 with confidence. The result of these experimental constraints is a paucity of information about
58 the number and identity of small proteins that are present even under standard growth conditions
59 in any biological system. Even less is known about stress-induced accumulation of these
60 proteins.

61 Those small proteins that have been identified and characterized, however, indicate that
62 they can have important roles in both bacteria and eukaryotes. In *Salmonella*, the 30 amino acid
63 MgtR protein negatively regulates the MgtC virulence factor by binding to and facilitating its
64 degradation (2). The 46 amino acid *Bacillus subtilis* Sda protein represses sporulation by
65 inhibiting the activity of the KinA kinase (4, 37). A group of 20-22 amino acid proteins that are
66 excreted by *Staphylococcus aureus* during infection act to disrupt neutrophil membranes and
67 cause cell lysis (48). In *Drosophila*, a family of 11 amino acid peptides encoded on a
68 polycistronic mRNA have been implicated in leg development (14), and the 31 amino acid

69 amphipathic helical sarcolipin protein from rabbit regulates the activity of the sarcoplasmic
70 reticulum Ca^{2+} -ATPase (49). These examples illustrate the diverse role of small proteins in cell
71 physiology, as well as their universal distribution.

72 In a previous study, we confirmed the synthesis of 20 previously-annotated small proteins
73 in *E. coli* by integrating the sequential peptide affinity tag (SPA) upstream of the stop codon on
74 the chromosome (22). 18 of the 20 proteins were expressed in rich media, while two were only
75 detected under specific conditions. We also predicted numerous new small protein genes using
76 sequence conservation and ribosome binding site models and confirmed the synthesis of 18 of
77 these proteins.

78 To further characterize 51 previously-detected or predicted proteins, we assayed the
79 levels of the SPA-tagged proteins under different growth conditions and after exposure to stress,
80 and found that a number of the small proteins are synthesized under specific growth conditions.
81 Four of these, YkgO, AzuC, YkgR and YobF, were selected for further investigation of the
82 mechanisms of their regulation. We found that transcription of the *ykgMO* operon is repressed
83 by Zur and that *azuC* and *ykgR* transcription is repressed and activated by Crp, respectively. We
84 also found that the increased accumulation of YobF in response to heat shock occurs at a post-
85 transcriptional level.
86

MATERIALS AND METHODS

Strain construction. All strains and oligonucleotides used in the study are listed in Tables S1 and S2, respectively. The SPA tagged strains constructed for this study were generated as previously described (22). The SPA fusion to the small open reading frame (ORF) predicted to overlap *pyrG* (*pyrG-mazG*) was generated in NM400 like for the other strains, but was not moved into MG1655. In this case, the NM400 strain was analyzed. The strains used for the dot blot assays retained the kanamycin cassette downstream of the SPA tag sequence to facilitate handling of the large number of cultures. Prior to the confirmation western blot assays, the kanamycin cassettes were removed by transformation with a plasmid expressing the FLP recombinase (pCP20) (7). Excision of the kanamycin cassette was confirmed by PCR. The $\Delta_{zur::kan}$ and $\Delta_{cadC::kan}$ mutants were generated by homologous recombination with a PCR fragment obtained by amplifying the kanamycin cassette on the pKD4 plasmid. The mutant alleles were sequenced, and the $\Delta_{zur::kan}$ allele was moved into the *ykgO*-SPA (kan^S) strain while the $\Delta_{cadC::kan}$ allele was moved into the *azuC*-SPA (kan^S) strain by P1 transduction. The $\Delta_{crp::cat}$ allele from NRD352 (10) and the *rpoS::Tn10* allele described in (52) were moved into the *azuC*-SPA (kan^S) and *ykgR*-SPA (kan^S) strains by P1 transduction. The $\Delta_{gadXYW::kan}$ allele from PM1293 (31) and the $\Delta_{gadE::kan}$ allele from EK551 (30) were moved into the *azuC*-SPA (kan^S) strain and the $\Delta_{oxyS::cat}$ allele from GSO113 (43) and the $\Delta_{lon::tet}$ allele from ML30008 (28) were moved into the *yobF*-SPA (kan^S) strain by P1 transduction.

Chromosomal transcriptional fusions between the *azuC* and *ykgR* promoters and the SPA tag were constructed by replacing the 5' untranslated region (UTR) of *azuC* and *ykgR* with the 5' UTR from multiple cloning site (MCS) 3 in the pBAD24 plasmid (5'-ACCCGTTTTTTGGGCTAACAGGAGGAATTAACC-3') (20) and the small protein ORF with

110 the SPA tag sequence. Chromosomal translational fusions to the SPA tag were constructed by
111 replacing the *azuC* and *ykgR* ORFs with the SPA tag sequence. The second codon in the SPA
112 tag sequence encodes Met, and this codon was used as the start codon for the SPA reporters. For
113 both sets of constructs, the SPA tag sequence and kanamycin cassette from the pJL148 plasmid
114 were amplified by PCR, transformed into NM400, sequenced and transduced by P1 into
115 MG1655. For the transcriptional and translational fusion strains, the kanamycin cassette was
116 retained in the strain after transduction.

117 The *ykgR*-SPA (kan^S) strain was transformed with pSAKT_{trc} from CAG62093 (B.-M.
118 Koo, unpublished) to overexpress σ^H under control of the *trc* promoter and with pTrc99A-*rpoE*
119 from KMT249 (K. M. Thompson, unpublished) to overexpress σ^E under control of the *trc*
120 promoter.

121 **Growth conditions.** All strains except those carrying *crp::cat* were grown in Luria Broth
122 (LB) rich medium or M63 minimal medium containing 0.0005% vitamin B1 and either 0.2%
123 glucose or 0.4% glycerol. For the *crp::cat* mutant strains, the minimal medium also contained
124 0.2% casamino acids.

125 For the dot blot assays, unless stated otherwise, all cultures were grown at 37°C as 5 ml-
126 cultures in 50 ml Falcon tubes with shaking at 250 rpm. LB cultures were inoculated with a
127 1:1000 dilution of overnight cultures. Minimal glucose cultures were inoculated with a 1:500
128 dilution of overnight cultures, while minimal glycerol cultures were inoculated with a 1:2000
129 dilution (to allow cultures to grow overnight). Oxygen-limited cultures were inoculated from
130 LB-grown overnight cultures into 2 ml LB + 0.2% glucose (deoxygenated) in 2 ml Eppendorf
131 tubes and were grown without shaking. The corresponding aerobic control cultures were
132 inoculated from the same overnight cultures into 5 ml LB + 0.2% glucose in 50 ml Falcon tubes

133 and were grown with shaking. Cells from both sets of samples were harvested at $OD_{600} = 0.4$ -
134 0.5 . To induce cell envelope stress, cultures grown to $OD_{600} = 0.2$ - 0.3 were exposed to 0.025%
135 SDS and 1 mM EDTA for 1 h . To induce acid stress, cultures grown overnight in LB were
136 inoculated into LB MES (100 mM , $\text{pH } 5.5$ - 5.6) or LB MOPS (100 mM , $\text{pH } 7.5$ - 7.6) media and
137 grown to $OD_{600} = 0.4$ - 0.5 . To induce heat shock, cultures grown at 30°C to $OD_{600} = 0.4$ - 0.5 were
138 transferred to a 45°C water bath for 20 min . To induce cold stress, cultures grown at 37°C to
139 $OD_{600} = 0.2$ - 0.3 were transferred to a 10°C water bath for 1 h . To induce oxidative stress or thiol
140 stress, cultures grown to $OD_{600} = 0.2$ - 0.3 were exposed to either 1 mM hydrogen peroxide or 1
141 mM diamide, respectively, for 30 min . For iron depletion stress, cultures grown to $OD_{600} = 0.1$ -
142 0.2 were treated with $200\text{ }\mu\text{M}$ dipyrldyl for 1 h . To induce DNA damage, cultures grown to
143 $OD_{600} = 0.3$ - 0.4 were treated with $2\text{ }\mu\text{g/ml}$ mitomycin C for 30 min . OD_{600} was measured at time
144 of harvest, and the control sample for each of the stress conditions other than oxygen-limitation
145 and acid stress were samples grown in LB medium to a similar OD_{600} . Cultures were transferred
146 to ice water baths in order to stop cell growth, and cells were collected from one ml of cells. The
147 cell pellets were frozen on dry ice and stored at -80°C .

148 Larger samples were used for the confirmation western blots. For the cell envelope stress
149 western blots, 30 ml -LB cultures started with a $1:1000$ dilution from overnight culture grown at
150 37°C were grown to $OD_{600} = 0.3$ - 0.4 at 37°C . The cultures were then divided into four 5-ml
151 aliquots and water, SDS (0.025%) and/or 1 mM EDTA (1 mM) was added to each culture.
152 Cultures were allowed to grow to $OD_{600} = 1.0$ - 1.2 (1 h of growth for all cultures except those
153 exposed to EDTA and SDS, which grew more slowly) and then harvested. For the acid stress
154 western blots, 30 ml -LB MOPS ($\text{pH } 7.5$) and LB MES ($\text{pH } 5.5$) cultures were inoculated $1:1000$
155 from overnight LB MOPS and LB MES cultures, and grown until $OD_{600} = 0.4$ - 0.5 at 37°C and

156 harvested. For the heat shock confirmation western blots, 30 ml LB cultures were inoculated
157 1:1000 from overnight cultures grown at 30°C, and were grown in 250 ml flasks to OD₆₀₀ = 0.2-
158 0.3 at 30°C. Then 10 ml of the culture was transferred to a 125 ml flask and incubated at 30°C
159 for 20 min, while another 10 ml of culture was incubated at 45°C for 20 min.

160 **Immunoblot assays.** For the dot blot assays, whole cells were resuspended in 1X sample
161 buffer [0.5% SDS, 0.006 mg bromophenol blue, 13% glycerol, 50 mM sodium phosphate buffer
162 (pH 8)] and heated at 95°C for 10 min. A 1-μl aliquot of each sample (equivalent to the cells in
163 OD₆₀₀ = 0.0028) was spotted on a nitrocellulose membrane (Invitrogen) followed by 1-μl
164 aliquots of a half dilution series. For cells expressing the YbgT-SPA, YnhF-SPA and YbhT-SPA
165 fusions, the levels of the proteins were significantly higher than the other tagged proteins, so the
166 boiled extract was diluted 1:10 in 1X sample buffer prior to spotting on the membrane. After all
167 spots were applied, the membrane was dried at room temperature for 10 min, and then incubated
168 in PBS-T (KD Medical) for 30 min. The membranes were blocked with 3% milk, and probed
169 with anti-FLAG M2-AP monoclonal antibody (Sigma-Aldrich) in 2% milk. Signals were
170 visualized using Lumi-Phos WB (Pierce) following standard methodologies. For the
171 confirmation western blots, samples were processed as previously described (9). In brief, a
172 fraction equivalent to the cells in OD₆₀₀ = 0.057 was separated on a Novex 16% Tricine gel
173 (Invitrogen) and transferred to a nitrocellulose membrane (Invitrogen). The membrane was
174 blocked with 3% milk and then probed with anti-FLAG M2-AP antibody (1:1000 dilution).
175 Membranes were then washed with PBS-T and incubated with Lumi-Phos (Pierce) prior to
176 exposure to film.

177 **Quantification of Protein Levels.** For quantification of western blot samples, dot blot
178 dilution series were used to determine relative protein levels. The dilution series were performed

179 similar to the dilution series for the survey dot blots; 1 µl-aliquots of a half dilution series were
180 spotted on nitrocellulose membranes adjacent to the dilution series of other samples in the
181 original western blot. Comparison of spot intensities of the different dilutions series allowed us
182 to quantify the relative protein levels for each sample.

183 **Northern analysis.** Northern analysis was conducted essentially as described (12).
184 Total RNA was collected from 5 ml of culture by acid-phenol extraction. RNA (5 µg) was
185 separated on 6% acrylamide gels, transferred to a Zeta-Probe Membrane (Bio-Rad) and probed
186 with oligonucleotide probes (listed in Table S2) end-labeled with ³²P-ATP using T4
187 polynucleotide kinase. Hybridization and wash steps were as described previously.

188 **Primer extension analysis.** Primer extension analysis was conducted as described (52).
189 Total RNA was collected from 5 ml of culture by TRIzol (Invitrogen) extraction.
190 Oligonucleotide primers (listed in Table S2) end-labeled with ³²P-ATP using T4 polynucleotide
191 kinase, were incubated with 5 µg of total RNA, allowed to anneal and then extended with reverse
192 transcriptase (Life Sciences). To generate a sequence ladder, a PCR fragment encompassing the
193 corresponding region was generated, and DNA sequencing reactions were carried out using the
194 SequiTherm EXCEL™ II DNA sequencing Kit (Epicentre) and the primer used in the primer
195 extension reaction.

RESULTS

196
197 **Small protein levels can be assayed in a high throughput manner.** We previously
198 confirmed the synthesis of 38 proteins of less than 50 amino acids by integrating the sequential
199 peptide affinity tag (SPA) upstream of the stop codon on the chromosome and assaying for
200 accumulation by western blot analysis (22). In our initial work we noted that some of the tagged
201 proteins were expressed under very specific conditions. For example we observed the expected
202 α -methylglucoside induction of SgrT (45). These observations suggested that the SPA tag did
203 not interfere with the regulated synthesis of the proteins. We thus decided to examine the levels
204 of the tagged proteins under a large range of conditions. We were especially interested in
205 determining if, under specific growth conditions, we could detect synthesis of full-length
206 proteins that we had not observed in cells grown in rich media in our previous study. These
207 predicted proteins included five previously annotated ORFs (Tpr, YlcH, DinQ, YoaI, YjjY) and
208 six ORFs predicted by our bioinformatic assays in the *ymjC'-ycjY*, *ycgI'-minE*, *ykgD-ykgE*, *gmr-*
209 *rnb*, *ydjA-sppA*, and *fabG-acpP* intergenic regions (Table S3). We also integrated the SPA tag
210 upstream of the stop codon of two additional putative small ORFs: *ymjB*, which was originally
211 annotated as a small ORF but is likely a pseudogene remnant, and a small ORF predicted to
212 overlap the 5' end of the *pyrG* ORF (Table S3).

213 Due to the high specificity and sensitivity of the α -FLAG antibody, it was possible to
214 screen accumulation of the 51 proteins mentioned above in a high-throughput manner using dot
215 blots (Fig. 1, Fig. S1 and Table 1). Based on dilution series of control samples, the dot blot
216 assays provided a detection range of greater than 1000 fold, depending on the exposure of the
217 blots. The large dynamic range allowed us to detect changes in protein accumulation of both the
218 highly expressed small proteins as well as small proteins expressed at much lower levels. To

219 reduce the number of false positives due to variations in the assay, only differences of 4-fold or
220 greater were considered significant. We did not detect the tagged proteins under any condition
221 for the 11 sORFs for which we had not observed synthesis in our previous study, and also not for
222 the two putative sORFs tagged specifically for this study. However, we did identify a number of
223 proteins that were highly induced under specific growth conditions; many of these proteins were
224 present at barely detectable levels under the previous conditions tested (22).

225 **Small protein levels vary with growth conditions.** To determine if any of the small
226 proteins were expressed differently in minimal medium compared to rich media, protein levels
227 were assayed during exponential phase growth in LB or M63 media supplemented with 0.2%
228 glucose or 0.4% glycerol. Five proteins (YccB-SPA, YncL-SPA, YkgO-SPA, YohP-SPA and
229 IlvX-SPA) were present at four-fold or higher levels in minimal glucose medium compared to
230 LB medium, while the levels of three proteins (YneM-SPA, YkgR-SPA and YoeI-SPA) were
231 lower (Fig. 1B). Similar changes in accumulation were observed for the comparison between
232 growth in glycerol minimal medium and LB, with some differences in the fold changes (Fig.
233 S1B). Consistent with the observed regulation, the IlvX protein is encoded in the *ilvXGEDA*
234 operon, which is induced in minimal media to synthesize isoleucine (42).

235 To determine if any of the small proteins were expressed differently depending on the
236 available carbon source, we examined protein levels in cells grown in minimal glucose versus
237 minimal glycerol medium (Fig. 1C). Four differences were detected. The YkgO-SPA, YnhF-
238 SPA and AzuC-SPA proteins were all present at lower levels in minimal glycerol-grown cells
239 compared to minimal glucose-grown cells while YkgR-SPA levels were higher in minimal
240 glycerol. The observed changes in protein levels could reflect regulation by CRP, a transcription
241 factor that modulates the expression of hundreds of genes depending on glucose availability (18).

242 CRP is able to bind DNA to positively or negatively regulate transcription after it complexes
243 with cyclic adenosine monophosphate (cAMP), the levels of which increase in response to low
244 glucose. Therefore, *ykgO*, *ynhF* and *azuC*, for which we observed higher protein levels in
245 glucose-grown cells, could be repressed by CRP-cAMP. In contrast, CRP could positively
246 regulate *ykgR*, for which lower protein levels were detected in glucose-grown cells. Potential
247 CRP binding sites can be found upstream of *azuC* and *ykgR* (see below).

248 Four proteins (YccB-SPA, YbgT-SPA, YnhF-SPA and YohP-SPA) showed more than 4-
249 fold higher levels during oxygen-limited growth compared to aerobic growth in rich media (Fig.
250 S1C). The most dramatic change was for the small protein YccB-SPA, with a greater than 30-
251 fold increase in protein levels. The *yccB* gene is in the *appCBA* operon, which encodes the
252 subunits of the cytochrome bd-II oxidase. The *ybgT* gene, a paralog of *yccB*, is in the *cydAB*
253 operon, which encodes cytochrome bd-I oxidase. Transcription of both of these operons has
254 been shown to be induced during low oxygen conditions (8, 9), and be activated by the ArcA
255 transcription factor (3, 8) (Table 1). Two proteins (YoeI-SPA and AzuC-SPA) showed reduced
256 levels under low oxygen conditions, suggesting that synthesis of these proteins is repressed
257 during oxygen-limited growth.

258 **Levels of small proteins increase upon stress.** The synthesis of a number of low
259 molecular weight proteins in *E. coli* is regulated upon exposure to stress. The IbpA (15.7 kD)
260 and IbpB (16.1 kD) proteins accumulate upon heat shock (25, 32), and the Csp family of proteins
261 (~7 kDa) accumulate in response to a variety of stresses, including cold shock (50, 51). To test
262 for stress-induced accumulation of our set of small proteins, we assayed the levels of the tagged
263 proteins after exposure to cell envelope stress, acidic pH, heat shock, cold shock, oxidative
264 stress, thiol stress, iron starvation and the DNA damaging agent mitomycin C. These stress

265 conditions were chosen to sample a broad range of potential stress response pathways and hence
266 possible small protein functions.

267 To examine small protein levels in cells undergoing cell envelope stress, cells were
268 exposed to SDS and EDTA for 1 h, and protein levels were compared to unstressed cells grown
269 to similar optical densities (Fig. 1D). Four proteins (YkgO-SPA, YneM-SPA, YohP-SPA and
270 YbgT-SPA) were induced at least 4-fold. Western blots confirmed that the YkgO-SPA and
271 YneM-SPA proteins accumulated after SDS/EDTA exposure (Fig. 2A). However, the levels of
272 YkgO-SPA and YneM-SPA were similarly increased when cells were exposed to EDTA alone,
273 suggesting that both proteins are likely induced in response to a decrease in cations due to
274 chelation by EDTA. Consistent with our results, *yneM* transcription was recently found to be
275 regulated by the two-component regulator PhoP in response to changes in Mg^{2+} levels (K. Moon
276 and S. Gottesman, unpublished). Increased levels of YohP-SPA appear to be specific to cells
277 exposed to both SDS and EDTA, suggesting that YohP is responding to cell envelope stress.
278 The dot blot results indicated that the YbgT-SPA protein is present at higher levels upon cell
279 envelope stress, but the levels of the protein were unchanged in the western blot assays.
280 Similarly according to the dot blots, YkgR-SPA levels were slightly decreased by SDS and
281 EDTA, but western blot analysis showed that while YkgR-SPA was present at lower levels after
282 SDS and/or EDTA exposure, the effect was slight. The differences between the dot blot and
283 western blot assays could be due to the absence of the kanamycin cassette in the tagged strains
284 used for the western blots, slight variations in growth, or differences between the two assays.

285 Two proteins (AzuC-SPA and YoaK-SPA) were present at higher levels under acidic (pH
286 5.5) compared to neutral (pH 7.5) conditions (Fig. 1E); AzuC-SPA was strongly induced, while
287 induction of YoaK-SPA was more modest (Fig. 2B). The 5' end of the transcript encoding *yoaK*

288 has not been mapped, but the gene may be encoded in a polycistronic message with the small
289 gene *yoaJ* as well as *yeaP*, a gene encoding a GGDEF domain protein with diguanylate cyclase
290 activity (39). Confirmation western blots showed that YoaJ, but not YoaK, seems to accumulate
291 to slightly higher levels in acidic media (Fig. 2B). Again the differences in the levels of the YoaJ
292 and YoaK proteins in the two assays could be due to the removal of the kanamycin cassettes
293 prior to the confirmation western blots, which could alter transcription or translation of other
294 genes encoded on the same mRNA. Further work is needed to determine if these two small
295 proteins are co-transcribed and coordinately regulated.

296 Eleven small proteins (YkgR-SPA, YobF-SPA, YqeL-SPA, YoaJ-SPA, YncL-SPA,
297 YpfM-Spa, YneM-SPA, YobI-SPA, AzuC-SPA, YthA-SPA and YccB-SPA) were found to
298 accumulate to at least four-fold higher levels after heat shock when cells were shifted from 30°C
299 to 45°C for 20 min (Fig. 1F). Western blot experiments confirmed the heat shock induction of
300 all of these proteins (Fig. 2C). However, the one protein that showed reduced levels after heat
301 shock, YbhT-SPA, did not show the same response in the confirmation western blot assays.
302 Nine proteins (YkgR-SPA, YobF-SPA, YqeL-SPA, YoaJ-SPA, YncL-SPA, YpfM-Spa, YneM-
303 SPA, YobI-SPA and AzuC-SPA) were induced within 5 min after transfer from 30°C to 45°C,
304 while two proteins (YthA-SPA and YccB-SPA) only showed induction after extended heat
305 shock. No σ^{32} binding sites have been predicted for any of these genes by RegulonDB (40)
306 (Table 1), though weak sites can be found upstream of *ypfM*, *yneM* and *yobI* (V. Rhodius,
307 unpublished). Four genes, *yobF*, *yqeL*, *ythA* and *yccB*, are predicted to be encoded in four
308 different operons, but the other genes in the operons have no clear relation to heat shock. Two of
309 the most strongly induced proteins, YkgR and YobF, were selected for further study (see below).
310 In contrast to heat shock, no proteins were induced more than two-fold by cold shock (Fig. S1D),

311 suggesting that none of the small proteins tested are involved in the cold shock stress response.
312 In fact, one heat shock induced protein, YkgR-SPA, was strongly repressed after cold shock,
313 suggesting that it may be detrimental during cold stress conditions.

314 To identify small proteins induced by oxidative stress, exponential phase cells were
315 exposed to hydrogen peroxide for 30 min and protein levels were compared with those in non-
316 stressed cells. Two proteins (YohP-SPA and AzuC-SPA) were induced ~4-fold by hydrogen
317 peroxide treatment, and YkgR-SPA levels were reduced ~4-fold (Fig. S1E). In a related
318 experiment, cells were exposed to the thiol oxidant diamide for 30 min (Fig. S1F). Diamide-
319 stressed cells showed increased levels of three proteins (AzuC-SPA, YkgR-SPA and YoaK-SPA)
320 as well as decreased levels of three proteins (YneM-SPA, YpdK-SPA and YoaJ-SPA).
321 Interestingly, there was little overlap in small protein accumulation when comparing oxidative
322 stress caused by hydrogen peroxide and thiol stress caused by diamide; only AzuC-SPA was
323 induced by both stresses. In contrast, the treatments had opposite effects on the levels of YkgR-
324 SPA.

325 The levels of the tagged proteins were essentially unchanged after 1 h of dipyrldyl
326 treatment to induce iron starvation (data not shown). Similarly, exposure to the DNA damaging
327 agent mitomycin C for 30 min did not significantly change the accumulation of any of the small
328 proteins (data not shown).

329 ***ykgM-ykgO* is repressed by Zur.** Five small proteins showed increased levels in cells
330 grown in minimal glucose medium compared to cells grown in rich medium. One of the most
331 highly induced proteins was YkgO, a paralog of the ribosomal protein RpmJ. We had observed
332 induction of YkgO-SPA in minimal media in our previous study, and hypothesized that this
333 effect was due to the regulation of the *ykgM-ykgO* operon by the Zur repressor (22). Zur, a

transcription factor found in many species of bacteria, regulates gene expression in response to zinc levels (34). DNA binding by Zur requires zinc; during zinc-depleted conditions, Zur is released from the DNA, allowing transcription of repressed genes. The minimal media used in our experiments does not contain added zinc and we also observed strong induction of YkgO-SPA when cells were treated with EDTA, consistent with our prediction that the high levels of YkgO-SPA in cells grown in minimal medium might be due to Zur derepression. A previous bioinformatic search for Zur binding sites in *E. coli* identified a potential binding site overlapping the transcription start of the *ykgM-ykgO* operon of *E. coli* (34) (Fig. 3A). To test whether the induction was in fact zinc-dependent, we added zinc to the media and found that *ykgM-ykgO-SPA* mRNA levels and YkgO-SPA protein levels were strongly repressed by zinc (Fig. 3B). In contrast, the levels of both the mRNA and protein were high in the presence of zinc in a Δzur background. These data confirm the predicted Zur repression of the *ykgM-ykgO* operon. We noted that YkgO-SPA levels were high in both exponential and stationary phase with the Δzur mutant, but only detected high levels of the protein in stationary phase *zur*⁺ cells grown in minimal media. A likely explanation for these observations is the presence of trace amounts of zinc in the minimal media used for these experiments; this level could be sufficient to repress *ykgM-ykgO* at low cell density but becomes limiting by stationary phase, allowing YkgO to be expressed in a majority of the cells. Consistent with this possibility, there was some variation in YkgO-SPA levels in exponential phase cells grown in minimal media, while YkgO-SPA levels were consistently high in stationary phase cells (data not shown).

***azuC* is repressed by CRP.** Three proteins were present at higher levels in minimal glucose medium as compared to minimal glycerol medium, suggesting that their synthesis may be repressed by the CRP transcription factor. One of these, AzuC, a basic (pKa = 10.3), 28

357 amino acid protein predicted to form an amphipathic α -helix by HeliQuest (15), is encoded by a
358 transcript that was first identified as the ISO92 small RNA (6). The start of the *azuC* transcript
359 was mapped to 42 nt upstream of the *azuC* AUG by primer extension (Fig. S2). Two locations
360 within the promoter region contain DNA sequences similar to CRP DNA binding sites (26); one
361 immediately upstream of the -35 hexamer and the other overlapping the -10 hexamer (Fig. 4A).
362 Binding of CRP to the putative downstream site would be expected to repress transcription
363 initiation. Consistent with this prediction, deletion of the *crp* gene eliminated the carbon source
364 regulation, resulting in similar levels of AzuC synthesis when cells were grown in media
365 containing glucose or glycerol (Fig. 4B).

366 In addition to being repressed by CRP, AzuC-SPA levels were repressed under low
367 oxygen conditions, moderately induced during heat shock, oxidative stress and thiol stress and
368 substantially elevated during growth in acidic medium. The strong acid induction is also
369 reflected in increased *azuC-SPA* mRNA levels (Fig. 5A). These observations suggested that
370 CRP repression is alleviated upon acid stress or that other DNA binding proteins regulate *azuC*
371 transcription. Alternatively, *azuC* mRNA levels might be subject to post-transcriptional
372 regulation. To test if *azuC* induction by low pH occurs at a transcriptional or post-transcriptional
373 level, we generated transcriptional and translational SPA fusions on the chromosome in which
374 the *azuC* ORF was replaced with the SPA tag sequence, beginning with an ATG at the second
375 codon of the tag sequence. For the transcriptional fusion, the *azuC* 5' UTR was also replaced
376 with the MCS 5' UTR from the pBAD24 plasmid. Although the relative levels of the SPA
377 peptide were different when expressed from different constructs, the levels from both *azuC*
378 constructs were elevated during growth in acidic media, but not from a control fusion to the *ykgR*
379 promoter (Fig. 5B). Quantification of relative protein levels of the fusions showed that the fold

380 difference between neutral and acidic-grown cells is ~8 fold for the full-length fusion, ~4 fold for
381 the transcriptional fusion and ~8 fold for the translational fusion (Fig S4A). These results
382 indicate that at least part of the acid induction occurs at both the transcriptional and post-
383 transcriptional levels.

384 To examine the contribution of CRP as well as known transcriptional regulators of the *E.*
385 *coli* responses to acidic conditions, we assayed AzuC-SPA levels in wild-type, Δcrp , $\Delta rpoS$,
386 $\Delta gadXW$, $\Delta gadE$ and $\Delta cadC$ cells exposed to pH 5.5 (Fig. 5C). No difference in the fold of
387 AzuC-SPA accumulation was detected between the wild type, $\Delta rpoS$, $\Delta gadXW$, $\Delta gadE$ and
388 $\Delta cadC$ cells (~30-fold). In contrast, the degree of AzuC-SPA induction in acidic media
389 compared to neutral media was reduced in a Δcrp background (~4-fold) (Fig S4B). These results
390 suggest that CRP regulation contributes to the pH-dependent regulation of AzuC levels, most
391 likely by repressing AzuC transcription in neutral media, but that there is also an as yet
392 unidentified transcriptional regulator of *azuC* induction in acidic medium. In contrast to the
393 strains carrying the full-length, transcriptional and translational fusions described above, the
394 kanamycin cassette associated with the *azuC-SPA* allele was flipped out in the strains assayed in
395 Fig. 5C and S4B, and we consistently observed higher fold changes for the AzuC-SPA fusion in
396 strains lacking the kanamycin cassette.

397 ***ykgR* is activated by CRP.** The dot blot assays showed that YkgR-SPA levels are higher
398 in media containing glycerol than glucose, suggesting that CRP could activate *ykgR*
399 transcription. Again primer extension analysis was used to map the start of the *ykgR* mRNA to a
400 single transcriptional start site 43 nt upstream of the *ykgR* AUG (Fig. S2). While a potential -10
401 sequence was noted, the promoter lacked a potential -35 sequence (Fig. 6A). However, a
402 potential CRP binding site could be found centered ~41 nucleotides upstream of the

transcriptional start, suggesting that the *ykgR* promoter may be a class II CRP-dependent promoter (27). At class II promoters, where the CRP binding site overlaps the -35 region and RNA polymerase makes DNA contacts upstream and downstream of CRP, there is generally a poor match to the consensus -35 sequence. Assays of YkgR-SPA synthesis in a Δcrp mutant confirmed that high levels of YkgR synthesis in minimal glycerol medium at 37°C are dependent on CRP (Fig. 6B). Although the *ykgR*-SPA transcript was barely detectable, the levels were higher for cells grown in minimal glycerol medium compared to minimal glucose medium and this difference was CRP-dependent (Fig. 6B).

In addition to being activated by CRP, YkgR-SPA accumulation was strongly elevated upon heat shock (Fig. 7A). The levels of the *ykgR*-SPA mRNA were similarly induced upon a shift to 45°C in the wild-type strain, though no transcript could be detected without or with heat shock in the Δcrp mutant strain (data not shown). To test if *ykgR* heat shock induction occurs at a transcriptional or post-transcriptional level, we generated SPA-fusion strains similar to those used to examine *azuC* expression. A transcriptional fusion was constructed in which the *ykgR* 5' UTR was replaced on the chromosome by the pBAD24 MCS 5' UTR and the *ykgR* ORF was replaced by the SPA tag sequence. For the translational fusion only the *ykgR* ORF was replaced by the SPA sequence. Heat shock led to similarly increased levels of the SPA tag from both fusions (Fig. 7B), suggesting that *ykgR* transcription is induced by heat shock. The genes of most heat shock-induced proteins are transcribed by the heat shock factor σ^{32} or the cell envelope stress regulator σ^E (1, 19). However, overexpression of neither σ^{32} nor σ^E from a plasmid led to YkgR-SPA induction (Fig. 7C). YkgR-SPA levels were also unchanged in $\Delta rseA$ cells that lack the σ^E anti-sigma factor and thus have elevated levels of σ^E (data not shown), as well as in an *rpoS* mutant strain (Fig. 7C). These data indicate that *ykgR* induction is not

426 regulated by these alternate σ factors, consistent with an absence of recognizable binding sites
427 upstream of the *ykgR* transcriptional start site. In addition, the results suggest that *ykgR* heat
428 shock induction is regulated by redundant or as-yet unidentified mechanisms.

429 **Heat shock induction of YobF occurs at a post-transcriptional level.** The levels of
430 the YobF-SPA protein were also strongly induced by heat shock (Fig. 2C). Primer extension
431 analysis showed that the *yobF-cspC* mRNA was transcribed from two different promoters
432 mapping 25-26 nt and 202 nt upstream of the *yobF* AUG (Fig. 8A, Fig. S2). However, in
433 contrast to *ykgR*, Northern analysis of the *yobF-cspC* operon after heat shock showed that
434 elevated temperature did not lead to increased levels of the mRNA (Fig. 8B). Together, these
435 data suggest that YobF heat shock induction occurs at a post-transcriptional level.

436 YobF-SPA accumulation after heat shock could be due to a decrease in degradation.
437 However, there were no changes in YobF-SPA levels in $\Delta clpP$, $\Delta clpQ$ and $\Delta clpY$ mutant strains
438 indicating that YobF is not a substrate for these proteases (data not shown). YobF is a substrate
439 for the Lon protease, but YobF-SPA levels still increase upon heat shock in a Δlon mutant (Fig.
440 8C). The possibility that YobF is a substrate of yet another protease cannot be ruled out.
441 Another plausible explanation for the observed heat shock induction is occlusion of the ribosome
442 binding site by mRNA secondary structure which is melted at higher temperatures as has been
443 observed for 5' UTRs that function as RNA thermometers (reviewed in reference (33)). The
444 *yobF-cspC* mRNA with either 5' end is predicted to be highly structured and this structure may
445 inhibit translation of the YobF protein (Fig. S3). In previous studies we noted a potential region
446 of basepairing between the hydrogen peroxide-induced OxyS small RNA and *yobF* in a region
447 overlapping the ribosome binding site (indicated in Fig. 8A) and showed that plasmid-expressed
448 OxyS led to significantly decreased *yobF* mRNA levels (44). We found that while endogenous

449 OxyS did not affect YobF-SPA levels under normal conditions (data not shown), the small RNA
450 did inhibit heat shock induction (Fig. 8D). If added immediately prior to heat shock, hydrogen
451 peroxide exposure reduced the level of YobF-SPA accumulation in an OxyS-dependent manner.
452 These data are consistent with the hypothesis that the 5'-UTR of the *yobF* transcript is generally
453 occluded in a secondary structure during normal growth, that this second structure unfolds upon
454 heat shock, and that heat shock-dependent unfolding of the *yobF-cspC* operon facilitates OxyS
455 repression of YobF translation. Hydrogen peroxide exposure is associated with a decrease in the
456 levels of *yobF-cspC* transcript in wild-type cells but not in the $\Delta oxyS$ mutant strain, consistent
457 with the assumption that OxyS promotes some degradation of the *yobF-cspC* mRNA.
458

DISCUSSION

459

460 The number of small proteins is not known for the proteome of any organism. A survey
461 of the literature shows the smallest proteins identified in most proteomics analyses in *E. coli* are
462 around 10 kDa (21), in contrast to the 2-5 kDa small proteins described in this work. In a
463 previous study, we showed that the *E. coli* K-12 genome encodes many more expressed small
464 proteins than had previously been predicted. Given that even less is known about the regulation
465 of small protein accumulation in *E. coli* and other organisms, we determined the levels of 51
466 confirmed and putative small proteins under a variety of growth conditions. We found many
467 show media- and stress-dependent regulation.

468 Of the 51 proteins tested, 21 were induced under at least one of the conditions tested.
469 Although these results show that the accumulation of many small proteins is subject to regulation
470 in response to environmental conditions, it is important to note possible limitations of these
471 experiments. Each of these small proteins is expressed as a chimeric protein with a C-terminal
472 SPA tag that is larger than the endogenous protein. The presence of this tag could potentially
473 inhibit normal regulation at the transcriptional, translational or post-translational level. The fact
474 that we observed previously predicted regulation, such as low oxygen induction of YbgT and
475 YccB, and observed the same regulation with the endogenous and tagged *ykgR* and *yobF-cspC*
476 transcripts (data not shown), suggest that the C-terminal SPA tag probably does not alter the
477 native regulation of most of the small proteins. It is also possible that some regulation may have
478 been missed because of the background levels of the SPA fragment we observed for a subset of
479 the proteins (22). High levels of this fragment could mask changes in levels of the full-length
480 protein in the initial dot blot assays. However, it is difficult to imagine a situation in which the
481 SPA tag leads to new regulation not associated with the native protein, and it is therefore likely

482 that the examples of regulation observed in this study accurately reflect that imposed on the
483 endogenous proteins.

484 **Transcription of some small protein genes is regulated by specific DNA binding**
485 **proteins.** Information about the transcriptional regulation of *E. coli* operons encoding some of
486 the small proteins is available. Of the small proteins examined in this study, four show
487 regulation consistent with previously predicted transcriptional regulation of the corresponding
488 operons. The synthesis of two small proteins, YbgT and YccB, which are encoded in
489 cytochrome oxidase operons with ArcA-dependent activation (3, 8), was induced by low oxygen
490 growth conditions as would be expected. Interestingly, the *appCB-yccB-appA* operon has been
491 suggested to be non-functional based on the inability of the operon to complement a cytochrome
492 oxidase mutant for aerobic growth on succinate-containing media (41). Rather than being
493 nonfunctional, the dramatic induction of YccB levels during growth in low oxygen (~30-fold)
494 compared to aerobic conditions suggests that the *appABC* cytochrome oxidase may only be
495 functioning during low oxygen conditions, in contrast to the two other cytochrome oxidases in *E.*
496 *coli*. The small protein IlvX, which is encoded in the *ilvXGMEDA* isoleucine biosynthesis
497 operon regulated by Lrp (35), shows the expected increase in protein levels in cells grown in
498 minimal media (~8 fold). Finally, the *ykgMO* operon was predicted to be regulated by the Zur
499 transcription factor (34); a prediction that was substantiated by the observed effects of zinc and a
500 Δ *zur* deletion. A recent microarray study of genes induced by zinc-depletion (17) identified
501 *ykgM* as a zinc-repressed gene, but did not report on *ykgO* as the microarray did not include
502 probes for the smaller gene. YtiA, a *Bacillus subtilis* ortholog of YkgM has also recently been
503 shown to be regulated by zinc through Zur (13).

504 In addition to confirming predicted regulation of small protein expression, we observed
505 new transcriptional regulation. We showed that CRP regulates the synthesis of at least two small
506 proteins: repressing *azuC* and activating *ykgR*. We also found that *azuC* transcription is induced
507 in acidic medium and *ykgR* transcription is induced by heat shock, but noted that previously
508 characterized transcriptional regulators of the acid and heat shock responses were not responsible
509 for the induction of these genes. Possibly there is redundancy in the transcriptional regulators
510 needed for the induction. Alternately, *azuC* and *ykgR* may be induced by other regulators that
511 were previously not associated with the acid and heat shock responses. A final possibility is that
512 CRP activity is altered under acid and heat shock stress conditions, and that this is responsible
513 for the changes in AzuC and YkgR levels. In fact, the difference between AzuC-SPA levels in
514 neutral and acidic media is reduced in the *crp* mutant. Little is known about the effect of acid
515 and heat stress on cAMP levels and CRP activity, but it has been shown that the expression of
516 other acid stress genes (*gadA* and *gadBC*) (5) and heat shock genes (*hslS*, *hslT*, and *htpG*) (16) is
517 altered in *crp* mutant strains.

518 **Stress induction of some small proteins occurs at a post-transcriptional level.** There
519 was little information about possible transcriptional regulation for small proteins not encoded in
520 operons or encoded in operons of unknown function. Many of these genes were only recently
521 annotated and would have been missed in microarray analyses carried out using gene-specific
522 probes. It is also possible that the accumulation of some proteins is regulated at the level of
523 translation or protein stability and thus would have been missed in the surveys of transcriptional
524 responses to stress. We obtained evidence that the heat shock induction of YobF occurs at the
525 post-transcriptional level. One potential mechanism for the post-transcriptional heat shock
526 induction of proteins is the melting of temperature-sensitive RNA stems that occlude the

527 ribosome binding site at normal temperatures (46). Consistent with this model, the *yobF-cspC*
528 mRNA leaders are predicted to fold into a secondary structure that could block ribosome binding
529 (Fig. S3). It is noteworthy that the predicted ribosome *yobF* binding site is unusually far
530 removed from the AUG (13 nucleotide spacer), another factor that could impact the heat shock
531 regulation of YobF synthesis. The findings that the *yobF-cspC* transcript is the target of the
532 OxyS small RNA and that YobF-SPA is a substrate of the Lon protease, support the conclusion
533 that YobF accumulation is subject to significant post-transcriptional control.

534 We suggest that still other small proteins are subject to post-transcriptional regulation.
535 M-fold predictions (53) of the 5'UTRs of at least one other heat shock induced protein, YqeL,
536 showed that this coding sequence may be preceded by structures reminiscent of a 4U-like motif,
537 a temperature-sensitive RNA structure where the ribosome binding site is paired with four
538 uridine nucleotides (47). If RNA melting contributes to YqeL induction and possibly other small
539 proteins, this would allow for a rapid increase in protein synthesis in response to the stress.
540 Another possibility is that other small heat shock-induced proteins besides YobF are protease
541 targets, and that they accumulate after heat shock because the proteases are being overwhelmed
542 with other unfolded protein substrates. An examination of the small protein levels in protease
543 mutants will be valuable for discerning whether this regulation contributes to the heat shock
544 induction.

545 **Possible roles of small proteins in stress responses.** Nineteen of the proteins that were
546 found to be induced under specific conditions are predicted to contain single transmembrane
547 helices, and many of these have been shown to fractionate with the cell membrane (22). For
548 these proteins, it is tempting to speculate that they could interact with the inner membrane under
549 conditions of slow growth or stress and thus modulate the function of other transmembrane

550 proteins, affect membrane permeability or serve a stabilizing role in the membrane. Other stress-
551 inducible proteins could be acting as chaperones or facilitate protein degradation upon stress
552 exposure; roles that have been determined for IbpA and IbpB (32). Yet another possibility is that
553 some of the proteins are of prophage origin. The heat shock-induced YthA protein is predicted
554 to be encoded in the *yjhB-yjhC-ythA* operon, which is flanked by IS elements and shows no
555 conservation outside of *E. coli*. Phage-related proteins often accumulate under stress conditions,
556 presumably so that the prophage can become lytic (36). One might imagine that small
557 hydrophobic proteins such as YthA could be descended from small transmembrane holin
558 proteins used by phage to lyse the bacteria host.

559 In a separate but complementary study, mutants carrying bar-coded deletions of the genes
560 encoding the small proteins were screened for sensitivity to cell envelope and acid stress in
561 large-scale competition experiments (23). Surprisingly, there was little overlap between the
562 regulation we observed and the phenotypes of the deletion strains. This discrepancy may be due
563 to the fact that the conditions used in the competition assays were different from those used for
564 the studies described here. Contrary to the short-term and relatively mild exposure to stress used
565 for these expression assays, the survival assays described by Hobbs et al. consisted of a long
566 term exposure to SDS/EDTA and treatment with extreme acid (pH 1.8). It is possible that the
567 stress-induced small proteins identified in the studies described here are not involved in more
568 severe stress conditions as tested in the competition studies. Alternatively, some of the small
569 proteins may have functions that are redundant with other stress response proteins. The limited
570 convergence between the two approaches highlights the importance of carrying out multiple
571 methodologies in characterizing these small proteins of unknown function.

572 The results presented here show that *E. coli* contains many stress-induced proteins that
573 were missed using classical biochemical techniques. Given the lack of probes for these genes on
574 most microarrays, the genes also were not assayed in whole genome expression surveys to
575 characterize the regulons of well-studied global transcriptional regulators such as Crp (16).
576 Based on these and other findings, it is clear that small proteins and the genes that encode them
577 need to be taken into consideration when designing future experiments. In addition, information
578 about the accumulation of the small proteins can set the stage for further functional
579 characterization, pointing to growth conditions best suited for biochemical assays such as co-
580 purification as well as for phenotypic assays of mutant strains.

ACKNOWLEDGMENTS

We thank S. Goodwin for conducting the primer extension analysis of the *yobF* operon, S. Gottesman, K. Moon and V. Rhodius for sharing unpublished data, the S. Gottesman, C. Gross and M. Maurizi labs for strains, A. Huerta for originally identifying the CRP binding site upstream of AzuC, and members of the Storz lab for helpful discussion and comments. We would particularly like to thank B.-M. Koo and E. Gogol for sharing the unpublished strain CAG62093 and K. M. Thompson and N. Majdalani for unpublished strain KMT249.

This research was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development and by postdoctoral fellowships from the Life Sciences Foundation (M.R.H.) and the National Research Council (B.J.P.).

REFERENCES

1. **Ades, S. E.** 2008. Regulation by destruction: design of the σ^E envelope stress response. *Curr. Opin. Microbiol.* **11**:535-540.
2. **Alix, E., and A. B. Blanc-Potard.** 2008. Peptide-assisted degradation of the *Salmonella* MgtC virulence factor. *EMBO J.* **27**:546-557.
3. **Brøndsted, L., and T. Atlung.** 1996. Effect of growth conditions on expression of the acid phosphatase (*cyx-appA*) operon and the *appY* gene, which encodes a transcriptional activator of *Escherichia coli*. *J. Bacteriol.* **178**:1556-1564.
4. **Burkholder, W. F., I. Kurtser, and A. D. Grossman.** 2001. Replication initiation proteins regulate a developmental checkpoint in *Bacillus subtilis*. *Cell* **104**:269-279.

- 604 5. **Castanie-Cornet, M. P., and J. W. Foster.** 2001. *Escherichia coli* acid resistance:
 605 cAMP receptor protein and a 20 bp cis-acting sequence control pH and stationary phase
 606 expression of the *gadA* and *gadBC* glutamate decarboxylase genes. *Microbiology*
 607 **147**:709-715.
- 608 6. **Chen, S., E. A. Lesnik, T. A. Hall, R. Sampath, R. H. Griffey, D. J. Ecker, and L. B.**
 609 **Blyn.** 2002. A bioinformatics based approach to discover small RNA genes in the
 610 *Escherichia coli* genome. *Biosystems* **65**:157-177.
- 611 7. **Cherepanov, P. P., and W. Wackernagel.** 1995. Gene disruption in *Escherichia coli*:
 612 Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-
 613 resistance determinant. *Gene* **158**:9-14.
- 614 8. **Cotter, P. A., V. Chepuri, R. B. Gennis, and R. P. Gunsalus.** 1990. Cytochrome o
 615 (*cyoABCDE*) and d (*cydAB*) oxidase gene expression in *Escherichia coli* is regulated by
 616 oxygen, pH, and the *fnr* gene product. *J. Bacteriol.* **172**:6333-6338.
- 617 9. **Dassa, J., H. Fsihi, C. Marck, M. Dion, M. Kieffer-Bontemps, and P. L. Boquet.**
 618 1991. A new oxygen-regulated operon in *Escherichia coli* comprises the genes for a
 619 putative third cytochrome oxidase and for pH 2.5 acid phosphatase (*appA*). *Mol. Gen.*
 620 *Genet.* **229**:341-352.
- 621 10. **De Lay, N., and S. Gottesman.** 2009. The Crp-activated small noncoding regulatory
 622 RNA CyaR (RyeE) links nutritional status to group behavior. *J. Bacteriol.* **191**:461-476.
- 623 11. **Fountoulakis, M., M. F. Takács, P. Berndt, H. Langen, and B. Takács.** 1999.
 624 Enrichment of low abundance proteins of *Escherichia coli* by hydroxyapatite
 625 chromatography. *Electrophoresis* **20**:2181-2195.

- 626 12. **Fozo, E. M., M. Kawano, Y. Kaya, K. Jones, K. E. Rudd, and G. Storz.** 2008.
 627 Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol.*
 628 *Microbiol.* **70**:1076-1093.
- 629 13. **Gabriel, S. E., and J. D. Helmann.** 2009. Contributions of Zur-controlled ribosomal
 630 proteins to growth under zinc starvation conditions. *J. Bacteriol.* **in press**.
- 631 14. **Galindo, M. I., J. I. Pueyo, S. Fouix, S. A. Bishop, and J. P. Couso.** 2007. Peptides
 632 encoded by short ORFs control development and define a new eukaryotic gene family.
 633 *PLoS Biol.* **5**:1052-1062.
- 634 15. **Gautier, R., D. Douguet, B. Antonny, and G. Drin.** 2008. HELIQUEST: a web server
 635 to screen sequences with specific α -helical properties. *Bioinformatics* **24**:2101-2102.
- 636 16. **Gosset, G., Z. Zhang, S. Nayyar, W. A. Cuevas, and J. Saier M.H.** 2004.
 637 Transcriptome analysis of Crp-dependent catabolite control of gene expression in
 638 *Escherichia coli*. *J. Bacteriol.* **186**:3516-3524.
- 639 17. **Graham, A. I., S. Hunt, S. L. Stokes, N. Bramall, J. Bunch, A. G. Cox, C. W.**
 640 **McLeod, and R. K. Poole.** 2009. Severe zinc depletion of *Escherichia coli*: Roles for
 641 high affinity zinc binding by ZinT, zinc transport and zinc-independent proteins. *J. Biol.*
 642 *Chem.* **284**:18377-18389.
- 643 18. **Grainger, D. C., D. Hurd, M. Harrison, J. Holdstock, and S. J. Busby.** 2005. Studies
 644 of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along
 645 the *E. coli* chromosome. *Proc. Natl. Acad. Sci. USA* **102**:17693-17698.
- 646 19. **Guisbert, E., T. Yura, V. A. Rhodius, and C. A. Gross.** 2008. Convergence of
 647 molecular, modeling, and systems approaches for an understanding of the *Escherichia*
 648 *coli* heat shock response. *Microbiol. Mol. Biol. Rev.* **72**:545-554.

- 649 20. **Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith.** 1995. Tight regulation,
650 modulation, and high-level expression by vectors containing the arabinose P_{BAD}
651 promoter. *J. Bacteriol.* **177**:4121-4130.
- 652 21. **Han, M. J., and S. Y. Lee.** 2006. The *Escherichia coli* proteome: past, present, and
653 future prospects. *Microbiol. Mol. Biol. Rev.* **70**:362-439.
- 654 22. **Hemm, M. R., B. J. Paul, T. D. Schneider, G. Storz, and K. E. Rudd.** 2008. Small
655 membrane proteins found by comparative genomics and ribosome binding site models.
656 *Mol. Microbiol.* **70**:1487-1501.
- 657 23. **Hobbs, E. C., J. L. Astarita, and G. Storz.** 2009. Small RNAs and small proteins
658 involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: analysis
659 of a bar-coded mutant collection. *J. Bacteriol.* **submitted**.
- 660 24. **Huerta, A. M., and J. Collado-Vides.** 2003. σ^{70} promoters in *Escherichia coli*: specific
661 transcription in dense regions of overlapping promoter-like signals. *J. Mol. Biol.*
662 **333**:261-278.
- 663 25. **Laskowska, E., A. Wawrzynów, and A. Taylor.** 1996. IbpA and IbpB, the new heat-
664 shock proteins, bind to endogenous *Escherichia coli* proteins aggregated intracellularly
665 by heat shock. *Biochimie* **78**:117-122.
- 666 26. **Law, E. C., N. J. Savery, and S. J. Busby.** 1999. Interactions between the *Escherichia*
667 *coli* cAMP receptor protein and the C-terminal domain of the alpha subunit of RNA
668 polymerase at class I promoters. *Biochem. J.* **337**:415-423.
- 669 27. **Lawson, C. L., D. Swigon, K. S. Murakami, S. A. Darst, H. M. Berman, and R. H.**
670 **Ebright.** 2004. Catabolite activator protein: DNA binding and transcription activation.
671 *Curr. Opin. Struct. Biol.* **14**:10-20.

- 672 28. **Lies, M., and M. R. Maurizi.** 2008. Turnover of endogenous SsrA-tagged proteins
673 mediated by ATP-dependent proteases in *Escherichia coli*. J. Biol. Chem. **283**:22918-
674 22929.
- 675 29. **Link, A. J., K. Robison, and G. M. Church.** 1997. Comparing the predicted and
676 observed properties of proteins encoded in the genome of *Escherichia coli* K-12.
677 Electrophoresis **18**:1259-1313.
- 678 30. **Ma, Z., S. Gong, H. Richard, D. L. Tucker, T. Conway, and J. W. Foster.** 2003.
679 GadE (YhiE) activates glutamate decarboxylase-dependent acid resistance in *Escherichia*
680 *coli* K-12. Mol. Microbiol. **49**:1309-1320.
- 681 31. **Mandin, P., and S. Gottesman.** 2009. A genetic approach for finding small RNAs
682 regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-
683 component system. Mol Microbiol. **72**:551-565.
- 684 32. **Matuszewska, M., D. Kuczyńska-Wiśnik, E. Laskowska, and K. Liberek.** 2005. The
685 small heat shock protein IbpA of *Escherichia coli* cooperates with IbpB in stabilization of
686 thermally aggregated proteins in a disaggregation competent state. J. Biol. Chem.
687 **280**:12292-12298.
- 688 33. **Narberhaus, F., T. Waldminghaus, and S. Chowdhury.** 2006. RNA thermometers.
689 FEMS Microbiol. Rev. **30**:3-16.
- 690 34. **Panina, E. M., A. A. Mironov, and M. S. Gelfand.** 2003. Comparative genomics of
691 bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of
692 ribosomal proteins. Proc. Natl. Acad. Sci. USA **100**:9912-9917.

- 693 35. **Rhee, K. Y., B. S. Parekh, and G. W. Hatfield.** 1996. Leucine-responsive regulatory
694 protein-DNA interactions in the leader region of the *ilvGMEDA* operon of *Escherichia*
695 *coli*. J. Biol. Chem. **271**:26499-26507.
- 696 36. **Rokney, A., O. Kobiler, A. Amir, D. L. Court, J. Stavans, S. Adhya, and A. B.**
697 **Oppenheim.** 2008. Host responses influence on the induction of lambda prophage. Mol
698 Microbiol. **68**:29-36.
- 699 37. **Rowland, S. L., W. F. Burkholder, K. A. Cunningham, M. W. Maciejewski, A. D.**
700 **Grossman, and G. F. King.** 2004. Structure and mechanism of action of Sda, an
701 inhibitor of the histidine kinases that regulate initiation of sporulation in *Bacillus subtilis*.
702 Mol. Cell **13**:689-701.
- 703 38. **Rudd, K. E., I. Humphery-Smith, V. C. Wasinger, and A. Bairoch.** 1998. Low
704 molecular weight proteins: a challenge for post-genomic research. Electrophoresis
705 **19**:536-544.
- 706 39. **Ryjenkov, D. A., M. Tarutina, O. V. Moskvina, and M. Gomelsky.** 2005. Cyclic
707 diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of
708 the GGDEF protein domain. J. Bacteriol. **187**:1792-1798.
- 709 40. **Salgado, H., S. Gama-Castro, M. Peralta-Gil, E. Díaz-Peredo, F. Sánchez-Solano, A.**
710 **Santos-Zavaleta, I. Martínez-Flores, V. Jiménez-Jacinto, C. Bonavides-Martínez, J.**
711 **Segura-Salazar, A. Martínez-Antonio, and J. Collado-Vides.** 2006. RegulonDB
712 (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon
713 organization, and growth conditions. Nucleic Acids Res. **34**:D394-D397.

- 714 41. **Sturr, M. G., T. A. Krulwich, and D. B. Hicks.** 1996. Purification of a cytochrome *bd*
715 terminal oxidase encoded by the *Escherichia coli* *app* locus from a Δ *cyo* Δ *cyd* strain
716 complemented by genes from *Bacillus firmus* OF4. J. Bacteriol. **178**:1742-1749.
- 717 42. **Subrahmanyam, C. S., J. D. Noti, and H. E. Umbarger.** 1980. Regulation of *ilvEDA*
718 expression occurs upstream of *ilvG* in *Escherichia coli*: additional evidence for an
719 *ilvGEDA* operon. J. Bacteriol. **144**:279-290.
- 720 43. **Tjaden, B., S. S. Goodwin, J. A. Opdyke, M. Guillier, D. X. Fu, S. Gottesman, and**
721 **G. Storz.** 2006. Target prediction for small, noncoding RNAs in bacteria. Nucleic Acids
722 Res. 2006 May 22;34(9):2791-802. **34**:2791-2802.
- 723 44. **Tjaden, B., S. S. Goodwin, J. A. Opdyke, M. Guillier, D. X. Fu, S. Gottesman, and**
724 **G. Storz.** 2006. Target prediction for small, noncoding RNAs in bacteria. Nucleic Acid
725 Res. **34**:2791-2802.
- 726 45. **Wadler, C. S., and C. K. Vanderpool.** 2007. A dual function for a bacterial small RNA:
727 SgrS performs base pairing-dependent regulation and encodes a functional polypeptide.
728 Proc. Natl. Acad. Sci. USA **104**:20454-20459.
- 729 46. **Waldminghaus, T., A. Fippinger, J. Alfsmann, and F. Narberhaus.** 2005. RNA
730 thermometers are common in α - and γ -proteobacteria. Biol. Chem. **386**:1279-1286.
- 731 47. **Waldminghaus, T., N. Heidrich, S. Brantl, and F. Narberhaus.** 2007. FourU: a novel
732 type of RNA thermometer in *Salmonella*. Mol Microbiol. **65**:413-424.
- 733 48. **Wang, R., K. R. Braughton, D. Kretschmer, T. H. Bach, S. Y. Queck, M. Li, A. D.**
734 **Kennedy, D. W. Dorward, S. J. Klebanoff, A. Peschel, F. R. DeLeo, and M. Otto.**
735 2007. Identification of novel cytolytic peptides as key virulence determinants for
736 community-associated MRSA. Nat. Med. **13**:1510-1514.

- 737 49. **Wawrzynow, A., J. L. Theibert, C. Murphy, I. Jona, A. Martonosi, and J. H.**
738 **Collins.** 1992. Sarcolipin, the "proteolipid" of skeletal muscle sarcoplasmic reticulum, is
739 a unique, amphipathic, 31-residue peptide. Arch. Biochem. Biophys. **298**:620-623.
- 740 50. **Xia, B., H. Ke, and M. Inouye.** 2001. Mol Microbiol. Acquirement of cold sensitivity by
741 quadruple deletion of the *cspA* family and its suppression by PNPase S1 domain in
742 *Escherichia coli*. Mol Microbiol. **40**:179-188.
- 743 51. **Yamanaka, K., W. Zheng, E. Crooke, Y. H. Wang, and M. Inouye.** 2001. CspD, a
744 novel DNA replication inhibitor induced during the stationary phase in *Escherichia coli*.
745 Mol. Microbiol. **39**:1572-1584.
- 746 52. **Zhang, A., S. Altuvia, A. Tiwari, L. Argaman, R. Hengge-Aronis, and G. Storz.**
747 1998. The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I)
748 protein. EMBO J. **17**:6061-6068.
- 749 53. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction.
750 Nucleic Acids Res. **31**:3406-3415.

751
752

FIGURE LEGENDS

FIG. 1. Dot blot analysis of small protein levels under different growth conditions. (A) Diagram of spotting pattern. Each tagged protein is assayed under two conditions; designated (+) or (-). A two-fold dilution series for YbgT-SPA from cells grown in LB is at the bottom of each blot, with a final dilution of 1:2048. Predicted, unannotated ORFs for which we had not previously seen full-length protein expression are designated ORF1-7: ORF1=*ymjC'*-*ycjY*, ORF2=*ycgI'*-*minE*, ORF3=*ykgD*-*ykgE*, ORF4=*gmr-rnb*, ORF5=*ydjA*-*sppA*, ORF6=*fabG*-*acpP*, ORF7=*pyrG*-*mazG* (Table S3). (B) Dot blot of protein levels in LB (-) versus minimal glucose media (+). (C) Dot blot of protein levels in minimal glucose (-) versus minimal glycerol media (+). (D) Dot blot of protein levels in LB (-) or LB + 0.025% SDS + 1 mM EDTA (+). (E) Dot blot of protein levels in LB MOPS (pH 7.5) (-) or LB MES (pH 5.5) (+). (F) Dot blot of protein levels in LB at 30°C (-) or after transfer to 45°C (+). A longer exposure of this dot blot is presented in Fig. S2 shows the changes in the less abundant small proteins after heat shock. In all cases, strains were grown as 5 ml cultures in 50 ml Falcon tubes. With the exception of the cultures exposed to heat shock, all cells were grown at 37°C. Cells were collected and samples were analyzed by dot blots as described in Materials and Methods. Proteins whose expression is induced under the + condition are boxed in blue, while proteins whose expression is reduced under the + condition are boxed in red.

FIG. 2. Western blot analysis of small protein levels. (A) Proteins induced by SDS and/or EDTA exposure. Cells grown overnight in LB were diluted into 30 ml of LB and grown to OD₆₀₀ = 0.2-0.3. Cultures were then split into five 5-ml aliquots and exposed to water, 0.025% SDS, 1 mM EDTA or 0.025% SDS and 1 mM EDTA. Cells were harvested before stress (T0)

777 and at $OD_{600} = 1.2-1.7$. (B) Proteins induced by acidic conditions. Cells grown overnight in LB
 778 were diluted into 30 ml of either LB MOPS (pH 7.5) or LB MES (pH 5.5) and harvested at
 779 $OD_{600} = 0.3-0.4$. (C) Proteins induced by heat shock. Cells grown overnight in LB at 30°C
 780 were diluted into 30 ml of LB and grown at 30°C to $OD_{600} = 0.4$. Cultures were then split into
 781 three 10-ml aliquots. One set of samples was transferred to 45°C while the other half was kept at
 782 30°C. Cells were harvested before transfer (T0) as well as 5 and 20 min after transfer. Western
 783 blot analysis using anti-FLAG, alkaline phosphatase-conjugated antibodies was carried out with
 784 whole-cell extracts harvested from the cultures above. Star (*) denotes the band corresponding
 785 to the fusion protein. Exposure times were optimized for each panel for visualization here;
 786 therefore, the signal intensity shown does not indicate relative abundance between proteins.

787

788 FIG. 3. Zur repression of *ykgO*. (A) Sequence of *ykgM-ykgO* promoter and coding region. The
 789 +1 of transcription (24) is denoted with an arrow. Potential σ^{70} -10 and -35 sequences are
 790 indicated in bold, and the predicted Zur binding site is boxed. (B) *ykgO-SPA* mRNA (top) and
 791 YkgO-SPA protein (bottom) levels in MG1655 and Δzur cells grown in minimal glucose
 792 medium with (+) or without (-) zinc. Overnight cultures were grown up in M63 containing 0.2%
 793 glucose and 100 μ M zinc acetate. Cells were washed twice in M63 containing 0.2% glucose and
 794 diluted into M63 glucose media lacking or containing 100 μ M zinc acetate. Cells were harvested
 795 at exponential (E) and stationary (S) phase and tested for *ykgO-SPA* expression and YkgO-SPA
 796 synthesis. For the Northern analysis, total RNA (5 μ g of each sample) was separated on a 6%
 797 acrylamide gel. RNA was transferred to nitrocellulose and probed with an end-labeled
 798 oligonucleotide complementary to the *ykgO* ORF. The band runs at ~600 nucleotides, consistent
 799 with the expected size of the *ykgM-ykgO-SPA* transcript. Western blot analysis was performed as

800 stated for Fig. 2. Star (*) denotes band corresponding to the full-length SPA-tagged YkgO
801 protein.

802

803 FIG. 4. CRP repression of *azuC*. (A) Sequence of the *azuC* promoter and coding region. The +1
804 of transcription (position 1986025 of the *E. coli* K-12 genome) is denoted with an arrow.
805 Potential σ^{70} -10 and -35 sequences are indicated in bold and the predicted CRP binding sites are
806 boxed. (B) *azuC*-SPA mRNA (top) and AzuC-SPA protein (bottom) levels in MG1655 and
807 Δcrp cells grown in minimal glucose (glu) and minimal glycerol (gly) medium supplemented
808 with 0.2% casamino acids and 0.0005% vitamin B1. M63 glucose and M63 glycerol cultures (30
809 ml inoculated with overnight cultures grown in the respective medium) were grown at 37°C to
810 OD₆₀₀ = 0.3-0.4. Northern analysis was performed as described in Fig. 3. The RNA band runs at
811 ~400 nucleotides, consistent with the expected size of the *azuC*-SPA transcript. Western blot
812 analysis was performed as stated for Fig. 2. Star (*) denotes band corresponding to the full-
813 length SPA-tagged AzuC protein.

814

815 FIG. 5. Acid induction of *azuC*. (A) *azuC* mRNA (top) and AzuC-SPA protein (bottom) levels
816 in MG1655 in LB, minimal glucose (glu) and minimal glycerol (gly) medium buffered at pH 7.6
817 or 5.6. The RNA band runs at ~400 nucleotides, consistent with the expected size of the *azuC*-
818 SPA transcript. Cells were diluted into the respective medium from overnight cultures grown in
819 LB. (B) Acid induction of *azuC* transcriptional and translational fusions. Extracts from strains
820 containing the SPA-tagged *azuC* allele (*azuC*-SPA), a transcriptional fusion to the *azuC*
821 promoter (P_{azuC}-5'+SPA), a translational fusion to the *azuC* promoter and 5' UTR (P+5'_{azuC}-
822 SPA) or a control transcriptional fusion to the *ykgR* promoter (P_{ykgR}-5'+SPA) were probed for

823 SPA expression in LB MOPS (pH 7.5) and LB MES (pH 5.5). In the transcriptional fusions, the
 824 *azuC* and *ykgR* 5' UTRs were replaced by the MCS 5' UTR from pBAD24 and the ORFs were
 825 replaced by the SPA tag. In the translational fusion, the *azuC* ORF was just replaced by the SPA
 826 tag (see Materials and Methods). Cultures grown overnight in LB were diluted into 10 ml LB
 827 MOPS (pH 7.6) or LB MES (pH 5.6), and cells were harvested at $OD_{600} = 0.45-0.6$. (C) AzuC-
 828 SPA expression in wild-type and mutant cells grown in neutral and acidic media. Cultures
 829 grown overnight in LB were diluted into 5 ml LB MOPS (pH 7.6) or LB MES (pH 5.6), and cells
 830 were harvested at $OD_{600} = 0.45-0.6$. Northern analysis was performed as described in Fig. 3.
 831 Western blot analysis was performed as stated for Fig. 2. A single star (*) denotes band
 832 corresponding to the full-length SPA-tagged AzuC protein, and a double star (**) denotes band
 833 corresponding to the SPA peptide.

834

835 FIG. 6. CRP activation of *ykgR*. (A) Sequence of the *ykgR* promoter and coding region. The
 836 +1 of transcription (position 312510 of the *E. coli* K-12 genome) is denoted with an arrow. A
 837 potential σ^{70} -10 sequence is indicated in bold and the predicted CRP binding site is boxed. (B)
 838 Primer extension analysis of *ykgR*-SPA mRNA (top) and western blot analysis of YkgR-SPA
 839 protein (bottom) levels in MG1655 and Δcrp cells grown in minimal glucose (glu) and minimal
 840 glycerol (gly) medium supplemented with 0.2% caseamino acids and 0.0005% vitamin B1. M63
 841 glucose and M63 glycerol cultures (inoculated from overnight cultures grown in the respective
 842 medium) were grown at 37°C to $OD_{600} = 0.3-0.4$. Primer extension assays were conducted using
 843 5 μ g total RNA of each sample and an end-labeled oligonucleotide complementary to the *ykgR*
 844 ORF. Western blot analysis was performed as stated for Fig. 2. Star (*) denotes band

845 corresponding to the full-length SPA-tagged YkgR protein.

846

847 FIG. 7. Heat shock induction of *ykgR*. (A) Primer extension analysis of *ykgR*-SPA mRNA (top)

848 and western blot analysis of YkgR-SPA protein (bottom) levels in MG1655 without and with

849 heat shock. 30-ml LB, M63 glucose or M63 glycerol cultures were inoculated a dilution of

850 overnight LB cultures and grown to $OD_{600} = 0.3-0.4$ before being split into three 10-ml aliquots.

851 Two aliquots were kept at 30°C while the other was incubated at 45°C. Cells were harvested

852 before transfer (T0) and after 20 min (30°C or 45°C). (B) Heat shock induction of *ykgR*

853 transcriptional and translational fusions. Extracts from strains containing the SPA-tagged *ykgR*

854 allele (*ykgR*-SPA), a transcriptional fusion to the *ykgR* promoter (P_{ykgR} -5'+SPA), a translational

855 fusion to the *ykgR* promoter and 5' UTR (P_{ykgR} -5'+SPA) or a control transcriptional fusion to the

856 *azuC* promoter (P_{azuC} -5'+SPA) were probed for SPA expression in cells without or with heat

857 shock. In the transcriptional fusions, the *ykgR* and *azuC* 5' UTRs were replaced by the MCS 5'

858 UTR from pBAD24 and the ORFs were replaced by the SPA tag. In the translational fusion, the

859 *ykgR* ORF was replaced by the SPA tag. Cultures grown overnight in LB were diluted into 30

860 ml LB and grown to $OD_{600} = 0.4-0.5$ before being split into two 10-ml aliquots. One aliquot was

861 kept at 30°C while the other was incubated at 45°C. (C) YkgR-SPA expression in wild-type cells

862 and cells with altered sigma factor levels. YkgR-SPA levels were assayed in wild-type and

863 $\Delta rpoS$ cells without and with heat shock, as well as in *ykgR*-SPA cells in which σ^H or σ^E

864 synthesis was induced by the addition of IPTG to half of the sample. MG1655 and $\Delta rpoS$ cells

865 grown overnight in LB were diluted into 30 ml LB and incubated at 30°C until $OD_{600} = 0.4$.

866 Cultures were then split into two 10-ml aliquots. One set of samples was transferred to 45°C

867 while the other was kept at 30°C. Cells were harvested before transfer (T0) and after 20 min

868 induction (30°C and 45°C). YkgR-SPA cells containing *rpoH* (pSAKTtrc) and *rpoE* (pCL245)
 869 overexpression plasmids were grown overnight in LB + 100 µg/ml carbenicillin and were diluted
 870 into 30 ml LB + 100 µg/ml carbenicillin. At OD₆₀₀ = 0.4, cultures were split into two 10-ml
 871 aliquots, and sigma factor expression was induced in one set of samples by adding IPTG to 1
 872 mM. Cells were harvested before induction (T0) and 20 min after induction (- and +). Primer
 873 extension assays were as conducted as described in Fig. 6. Western blot analysis was performed
 874 as stated for Fig. 2. A single star (*) denotes band corresponding to the full-length SPA-tagged
 875 YkgR protein, and a double star (**) denotes band corresponding to the SPA peptide.

877 FIG. 8. Post-transcriptional heat shock induction of YobF. (A) Sequence of the *yobF-cspC*
 878 promoter and coding region. The +1 of both the longer transcript (position 1905817 of the *E.*
 879 *coli* K12 genome) and the shorter transcript (position 1996212 of the *E. coli* K12 genome) are
 880 denoted by arrows. Possible σ^{70} binding sites are indicated in bold, and nucleotides that are
 881 predicted to base pair with the OxyS small RNA are denoted with dots. (B) *yobF-cspC* mRNA
 882 (top) and YobF-SPA protein (bottom) levels in cells exposed to heat shock. Samples were treated
 883 as in Fig. 7A. Cells were harvested before transfer (T0) and after 20 min at 30°C or 45°C. (C)
 884 YobF-SPA levels in a Δlon mutant strain. Again cell were harvested from samples kept at 30°C
 885 or 45°C for 20 min or before transfer (T0). (D) YobF heat shock induction with and without
 886 hydrogen peroxide exposure in MG1655 and an $\Delta oxyS$ mutant. Heat shock induction was
 887 conducted as described for (C), except that in some cases, hydrogen peroxide was added to the
 888 cells to a final concentration of 250 µM, 10 min prior to heat shock. Northern analysis was
 889 performed as described in Fig. 3. The prominent smaller RNA band runs at ~600 nucleotides,
 890 consistent with the expected size of the shorter *yobF-SPA* transcript. Western blot analysis was

891 performed as stated for Fig. 2. Star (*) denotes band corresponding to the full-length SPA-tagged
 892 YobF protein.

Table 1

Protein ^a	LB vs Minimal Glucose	Minimal Glucose vs Minimal Glycerol	Aerobic vs Low Oxygen	Mock vs SDS+EDTA	pH 7.5 vs pH 5.5	30°C vs 45°C	37°C vs 10°C	Mock vs H ₂ O ₂	Mock vs Diamide	Map ^b	RegulonDB ^c
YpFM						++				monocistronic	none
YccB	++ ^d		+++++			++				<i>appAB-yccB-appC</i>	σ^{38} , σ^{70} , AppY, ArcA
YncL	++					+++				monocistronic	none
YneM	- ^e			++		++			-	monocistronic	none
YbgT			+++	++						<i>cydAB-ybgTE</i>	σ^{70} , ArcA, Fnr, FruR, HNS
YkgO	+++++	-		+++++						<i>ykgMO</i>	σ^{70} , Zur
YqgB								-		<i>yqgB-speAB</i>	none
YnhF		-	++							monocistronic	none
YthA						++				<i>yjhBC-ythA</i>	none
YobI						++				monocistronic	NL
YqeL						++				<i>yqeKL</i>	none
YohP	++++		++++	++				++		monocistronic	NL
YkgR	-	++		-		+++++	-	-	++	monocistronic	NL
IlvX	+++									<i>ilvXGMEDA</i>	σ^{70} , Lrp, IHF
YoaJ						++			++	<i>yeaP-yoaKJ</i>	none
YpdK									-	monocistronic	NL
YoaK					++				-	<i>yeaP-yoaKJ</i>	none
YoeI	-		-							<i>yoeI-yeef</i>	none
YbhT						-				monocistronic	none
AzuC		-	-		++	+++		++	++	in IsrB sRNA	NL
YobF						++				<i>yobF-cspC</i>	none

^aAll small proteins that had at least a 4-fold change in protein levels in at least one experiment are listed

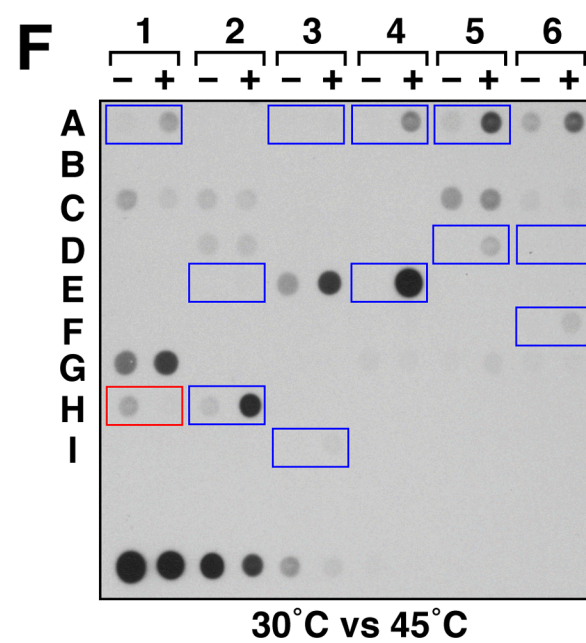
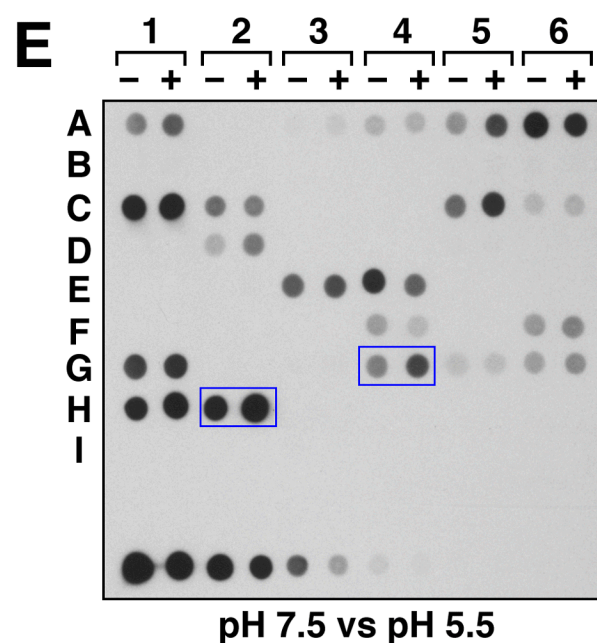
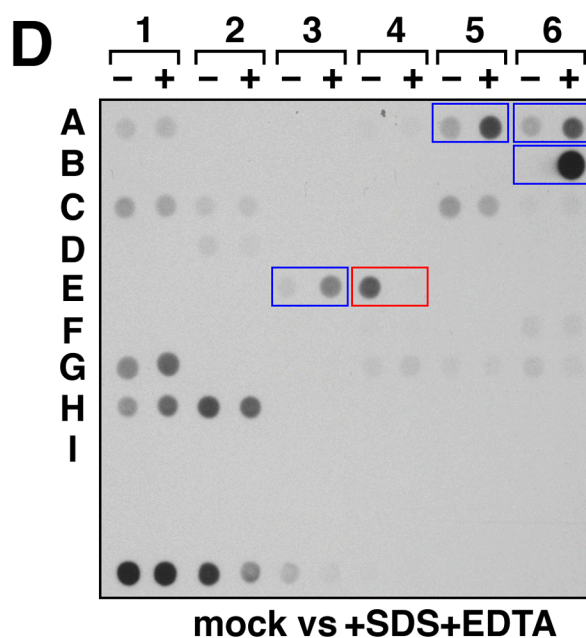
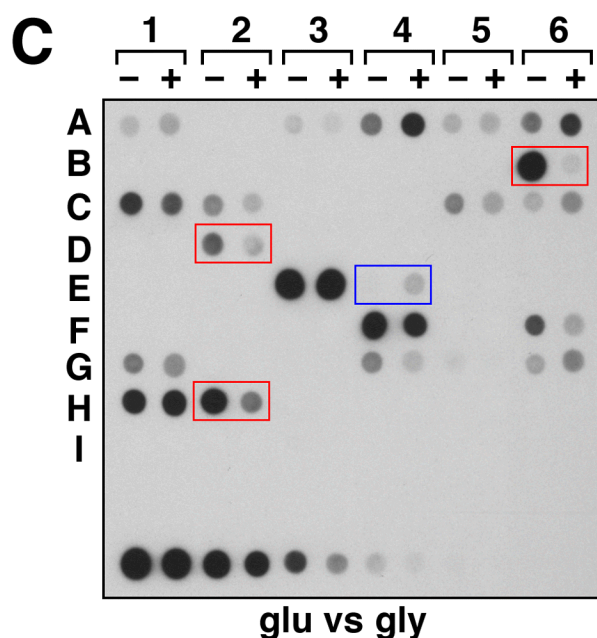
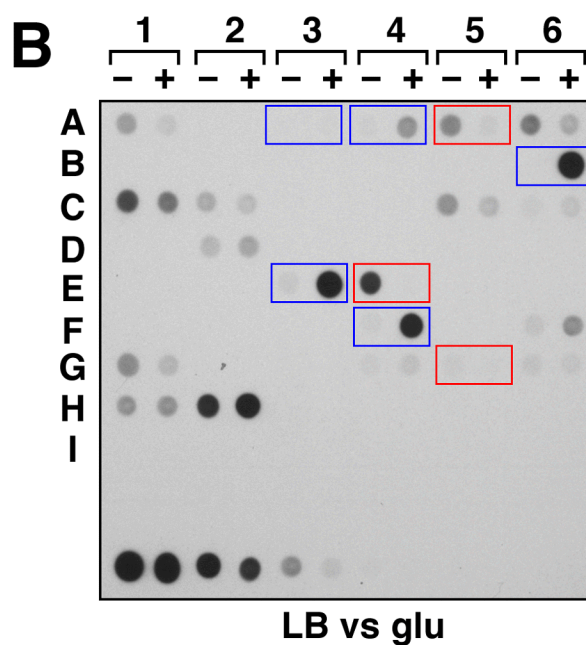
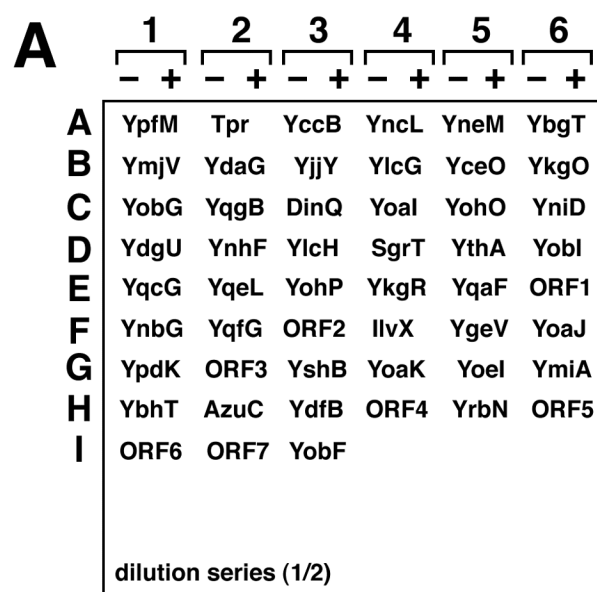
^bPredicted operonic organization of short ORFs

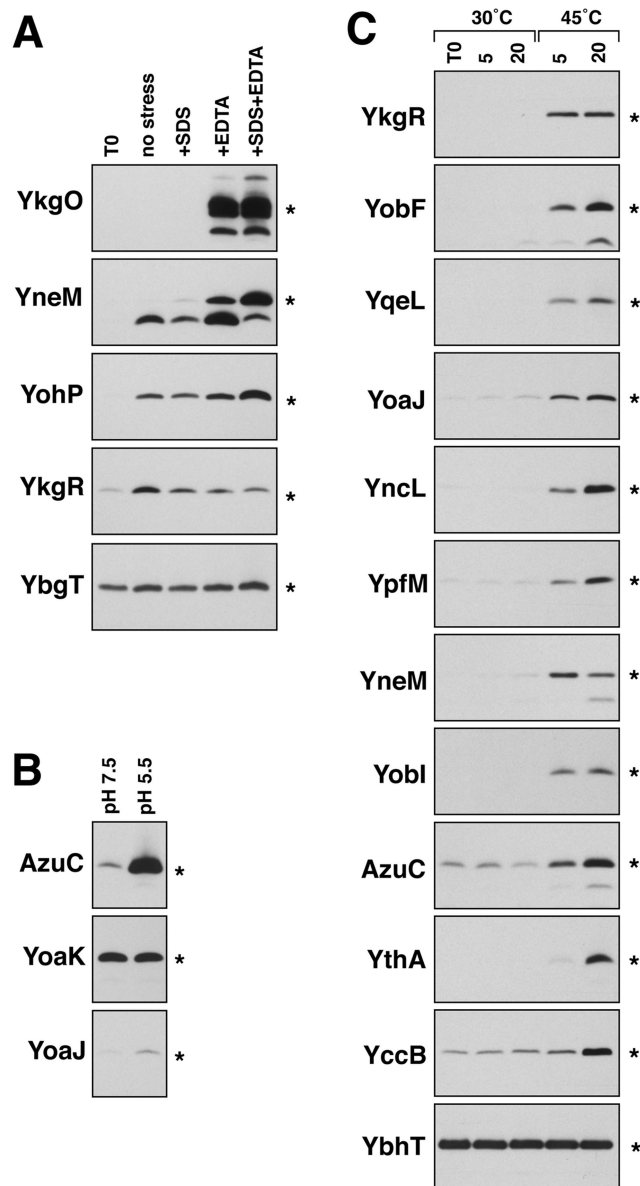
^cTranscriptional regulators listed in RegulonDB that are upstream of either the short ORF or the first gene in the operon; NL: not listed

^dEach (++) indicates a two-fold increase in protein expression in cells grown under the second condition compared to the first condition

^eEach (-) indicates a two-fold decrease in protein expression in cells grown under the second condition compared to the first condition

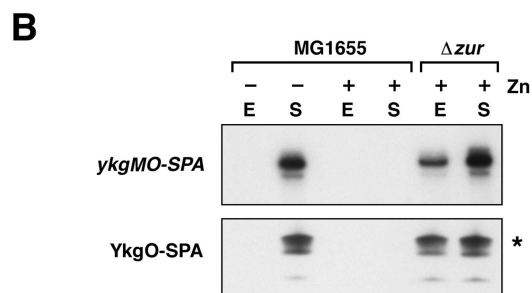
893





A

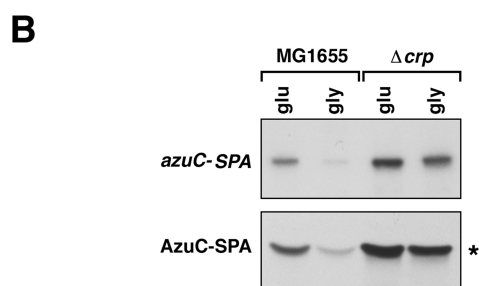
AAAATTCCAC ATCCACAAAG AGTCACAGGG ATTGAGTGTT **GAAATGATCC**
 GGATGAGCAT **G**TATCTTTAT **G**TTATGTTA TAACATAACA GGTA**AAAAATG**
 ↑
 ATGAAGCCCA ATATCCATCC TGAGTATCGT ACTGTGGTGT TCCACGACAC
 M K P N I H P E Y R T V V F H D T
 CAGTGTTGAT GAGTACTTTA AAATCGGCTC GACTATCAAA ACAGACCGTG
 S V D E Y F K I G S T I K T D R E
 AGATTGAGCT GGATGGCGTA ACGTATCCAT ACGTGACAAT TGATGTCTCT
 I E L D G V T Y P Y V T I D V S
 TCTAAATCGC ACCCGTTCTA TACAGGGAAG CTGAGAACAG TGGCATCAGA
 S K S H P F Y T G K L R T V A S E
 AGGAAATGTT GCACGATTCA CCCAACGTTT TGGTCGTTTT GTTAGCACGA
 G N V A R F T Q R F G R F V S T K
 AAAAGGGGGC GTGATGAAAG TTCTTAACTC TCTGCGTACC GCAAAAAGAAC
 K G A * M K V L N S L R T A K E R
 GCCATCCAGA CTGTCAGATT GTGAAGCGAA AAGGACGGCT ATATGTGATT
 H P D C Q I V K R K G R L Y V I
 TGTAAATCTA ATCCACGTTT TAAGGCCGTT CAGGGTCGTA AGAAAAAACG
 C K S N P R F K A V Q G R K K K R
 TTGATTCAAA ATTCGACGGA TTAACGATAT TTGTCTGATT AATAATCAGA
 *

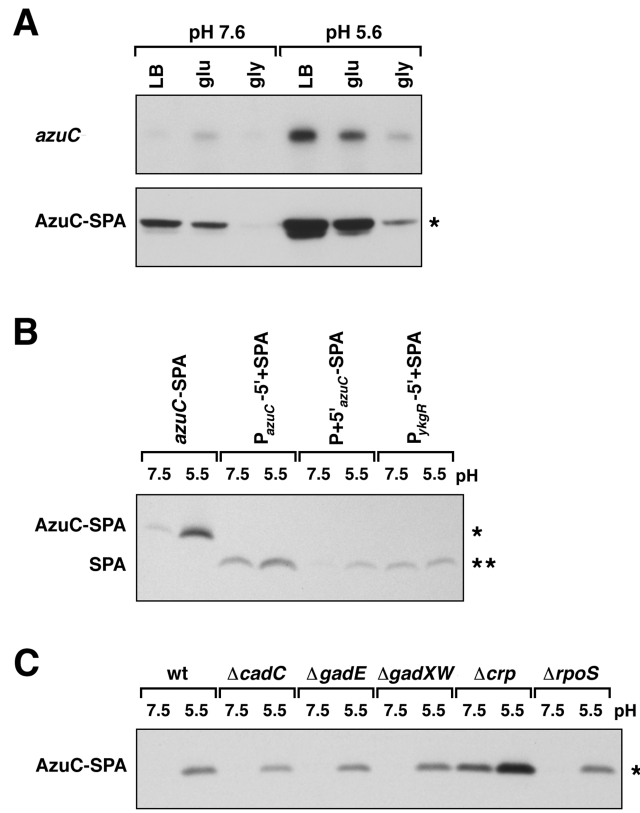


A

```

AACATTAAAG GGGTATTTAT GTGA TTTGTA TCACA TAAAG ATTGATTTT
GTGA TGGTTG TCATA TTATA AGCGCAAAGA CAATAACACC TGTATAACAA
      ↑
ATGGTCGGAG TGCCGCGATG AAACGCGCA AAATCCTGAA AAGTATGTTT
      M K L R K I L K S M F
AATAACTATT GCAAGACGTT CAAAGACGTA CCGCCAGGCA ATATGTTCCG
N N Y C K T F K D V P P G N M F R
ATAACAAAAA ACCTGCTCCG GCAGGTTTTT TTGTGTCCTG ATGACGGTGG
*
```





A

```

TTCAAATTG TGAACATGT CAAAAAATTG GTTGACCAA CTCGTTATTT
ATATAAGGGC ACTTACGAAG TGCACTCTTT TTTAAAGCGA GGAAGTACCA
      ↑
ATGAAAGAGA ATAAAGTACA GCAAATCAGT CATAAACTGA TTAATATCGT
M K E N K V Q Q I S H K L I N I V
TGTTTTGTG GCAATTGTAG AATACGCCTA TTTATTCTC CATTCTATT
V F V A I V E Y A Y L F L H F Y *
AATAACGGAA ATAACTGTT CACTTCAGTG ATATTAAAA TATGCATCCT

```

